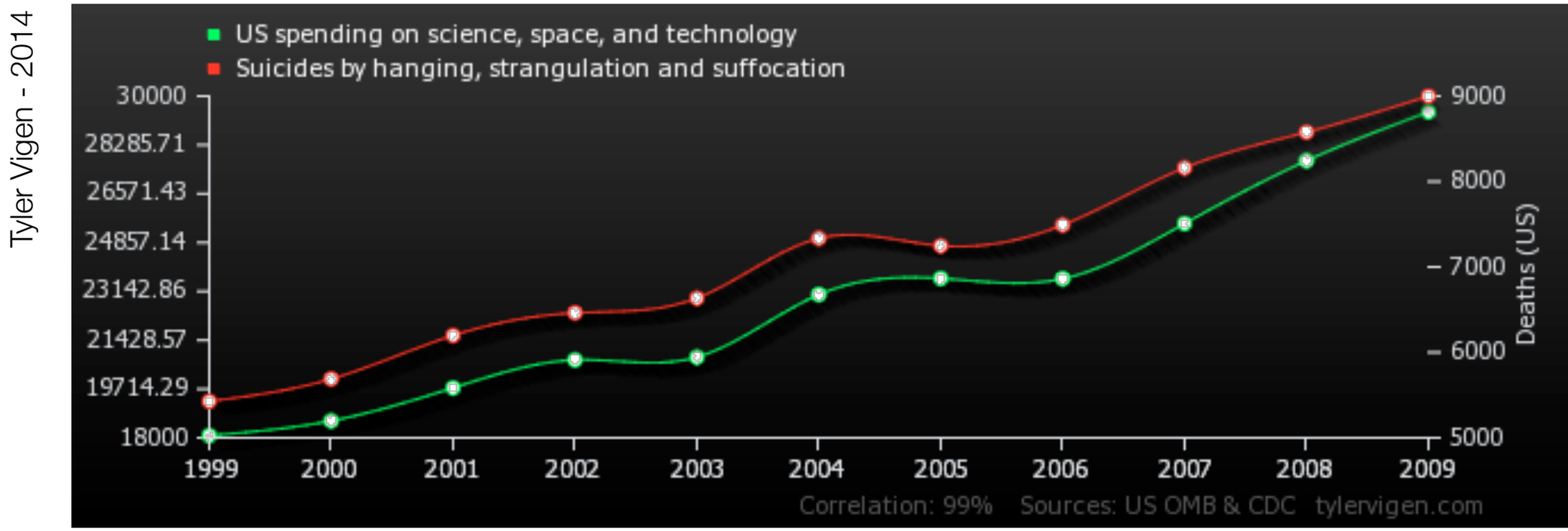
Big Data in Social Sciences

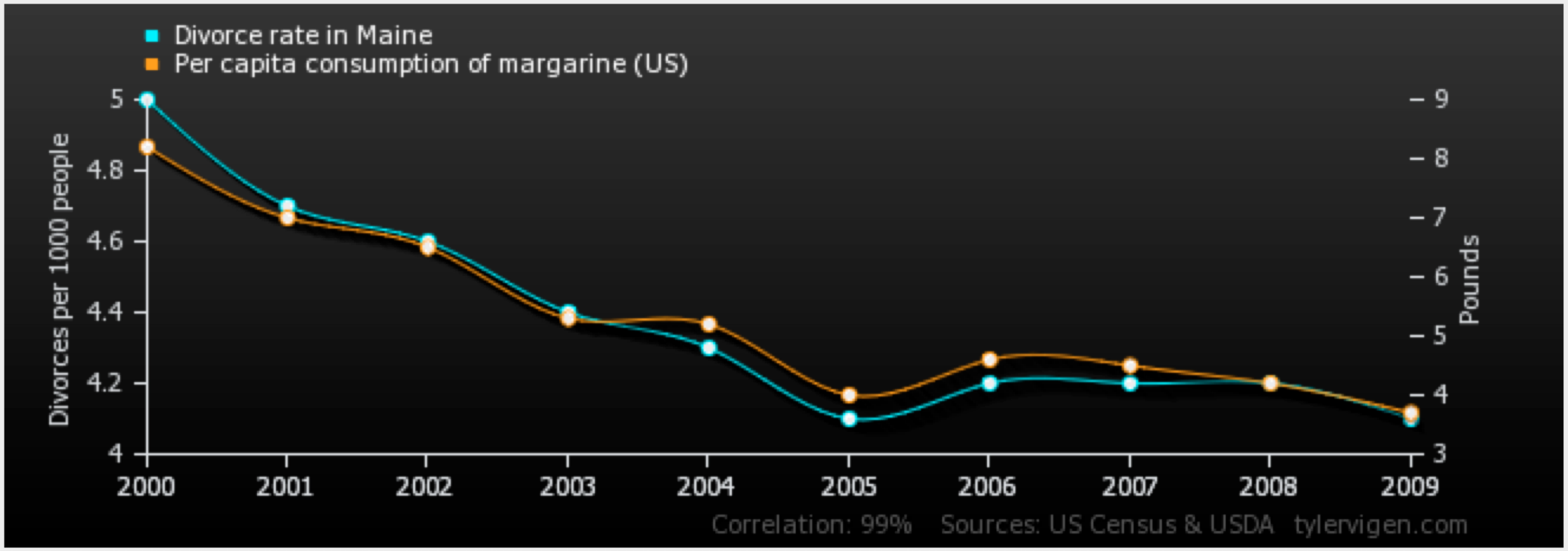


Boris Beaude - Chôros, École Polytechnique Fédérale de Lausanne

Divorce rate in Maine

correlates with

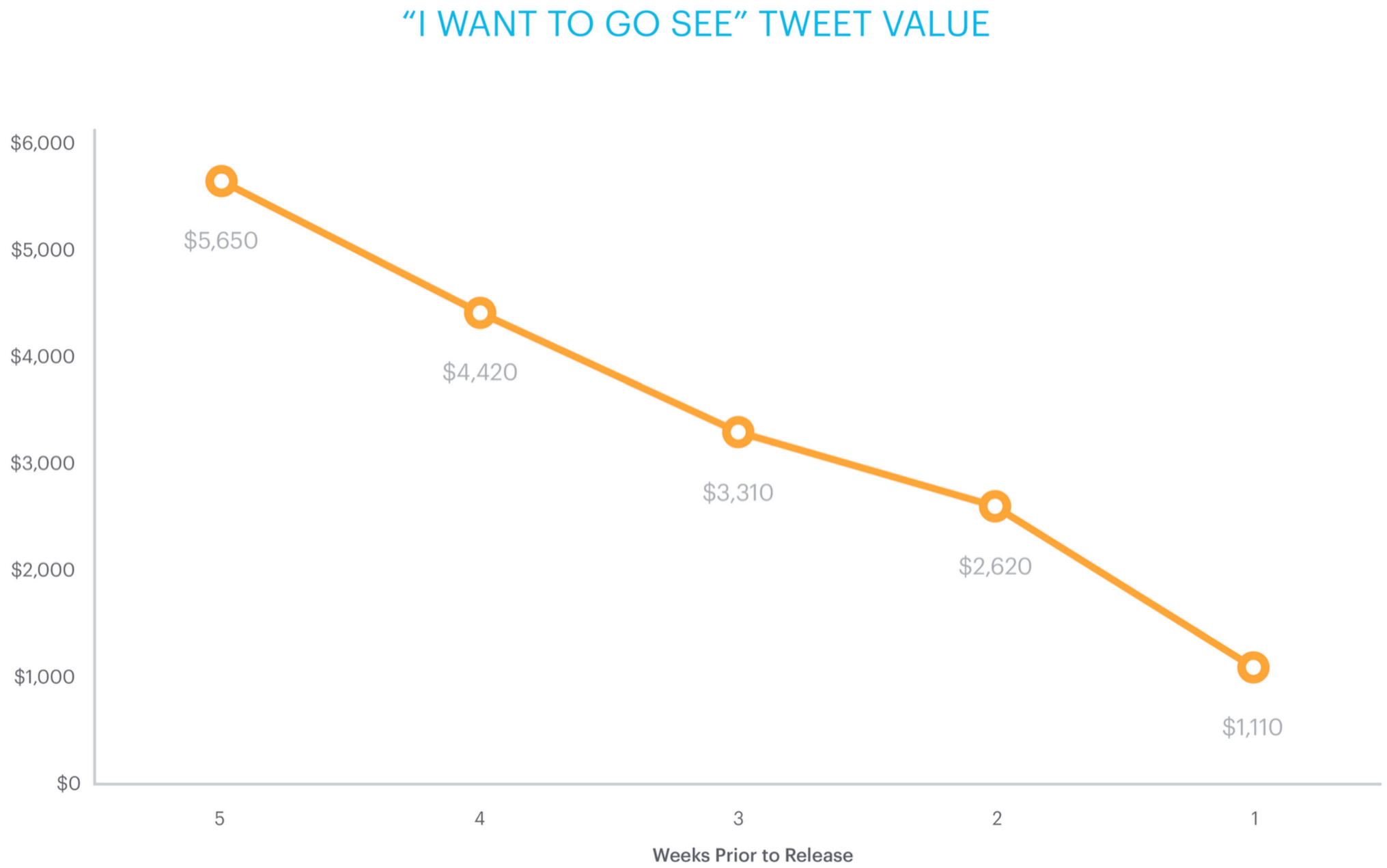
Per capita consumption of margarine (US)



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Divorce rate in Maine *Divorces per 1000 people (US Census)*	5	4.7	4.6	4.4	4.3	4.1	4.2	4.2	4.2	4.1
Per capita consumption of margarine (US) *Pounds (USDA)*	8.2	7	6.5	5.3	5.2	4	4.6	4.5	4.2	3.7

Correlation: 0.992558

Big Data in Social Sciences



How Much is a Single Tweet Worth? For Movies, 560 - Jaime Brugueras - 2015 - Networked Insights

Big Data in Social Sciences

So what should movie marketers take away from the data? Tweets matter, and the more direct and far in advance they are, the better.

Encourage fans to tweet about movies long before opening weekend or to make plans over Twitter to attend a showing with friends. Any way you can encourage an authentic conversation on social media about your film will expand your audience and raise opening weekend revenue as a result.

How Much is a Single Tweet Worth? For Movies, 560 - Jaime Brugueras - 2015 - Networked Insights

Big Data in Social Sciences

“An Interview with Marshall McLuhan”

Playboy, Mars 1969.

“The computer can be used to direct a network of global thermostats to pattern life in ways that will optimize human awareness”

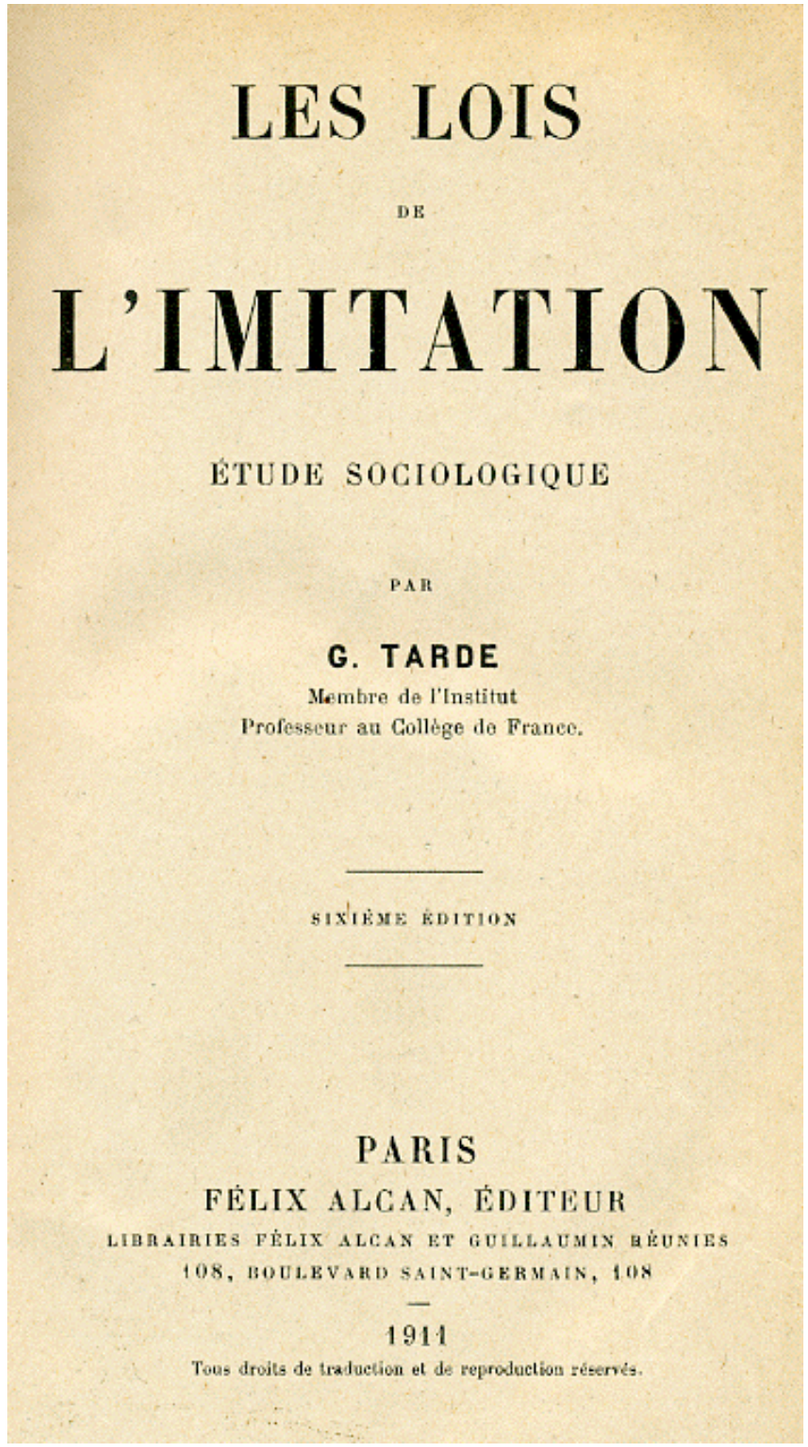
“Already, it’s technologically feasible to employ the computer to program societies in beneficial ways.”

Big Data in Social Sciences

« Les Lois de l'imitation »

Gabriel Tarde, 1890.

« Si la statistique continue à faire les progrès qu'elle a faits depuis plusieurs années, si les informations qu'elle nous fournit vont se perfectionnant, s'accélérant, se régularisant, se multipliant toujours, il pourra venir un moment où, de chaque fait social en train de s'accomplir, il s'échappera pour ainsi dire automatiquement un chiffre, lequel ira immédiatement prendre son rang sur les registres de la statistique, continuellement communiquée au public et répandue en dessins par la presse quotidienne ». p. 192


LES LOIS

DE

L'IMITATION

ÉTUDE SOCIOLOGIQUE

PAR

G. TARDE

Membre de l'Institut

Professeur au Collège de France.

SIXIÈME ÉDITION

PARIS

FÉLIX ALCAN, ÉDITEUR

LIBRAIRIES FÉLIX ALCAN ET GUILLAUMIN RÉUNIES

108, BOULEVARD SAINT-GERMAIN, 108

1911

Tous droits de traduction et de reproduction réservés.

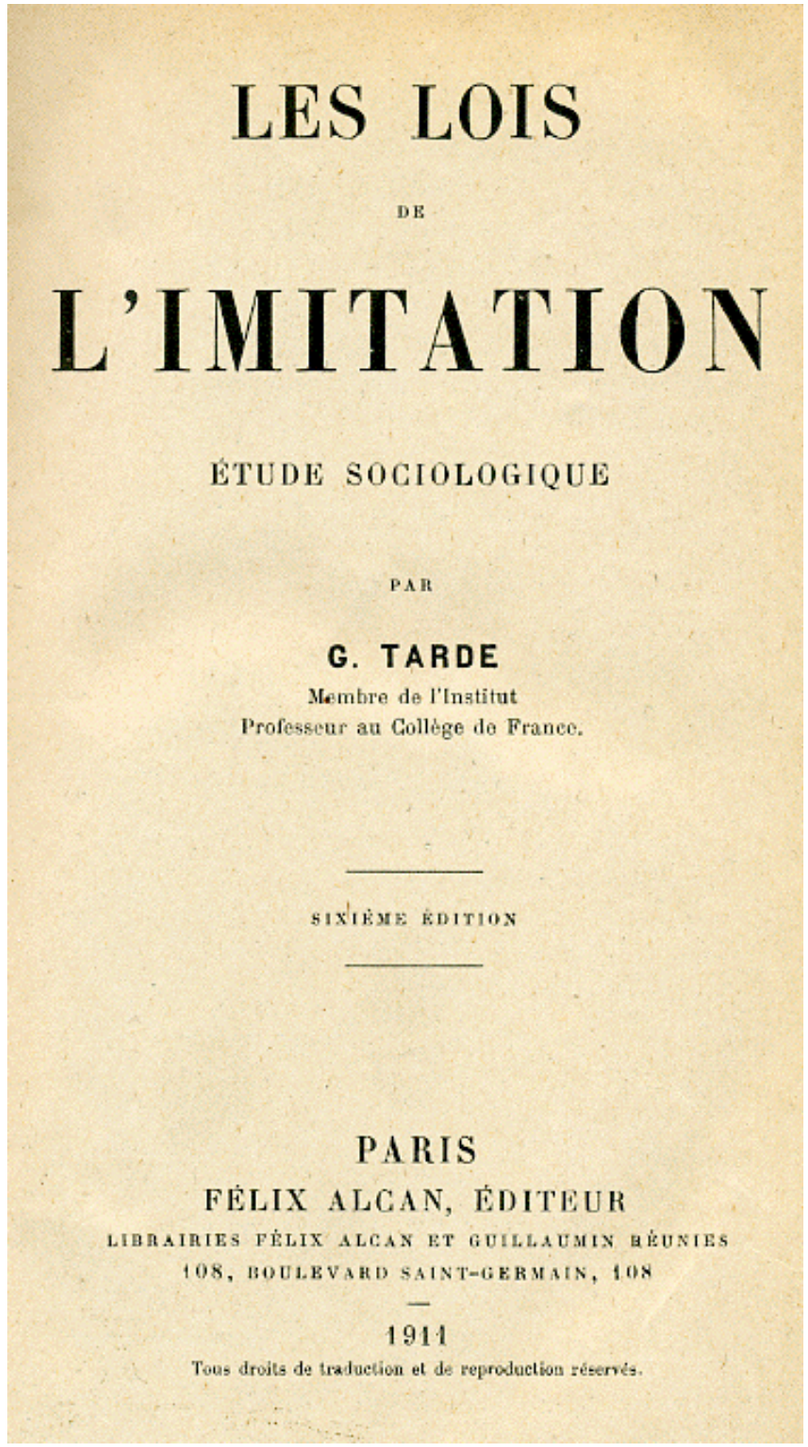
« Les Lois de l'imitation »

Gabriel Tarde, 1890.

« Est-ce qu'on ne surfait pas la statistique quand on émet, à propos d'elle, certaine espérance qu'il me faut indiquer en finissant ?

Comme on voit ses résultats numériques se régulariser, affecter plus de constance, à mesure qu'elle porte sur de plus grands nombres, on est quelquefois enclin à penser que, bien plus tard, si la marée montante de la population continue à accroître et les grands États à grandir, un moment viendra où tout, dans les phénomènes sociaux, sera réductible en formules mathématiques. D'où l'on induit abusivement que le statisticien pourra un jour prédire l'état social futur aussi sûrement que l'astronome la prochaine éclipse de Vénus. En sorte que la statistique serait destinée à plonger toujours plus avant dans l'avenir comme l'archéologie dans le passé.

Mais nous savons par tout ce qui précède que la statistique est circonscrite dans le champ de l'imitation et que celui de l'invention lui est interdit ». p. 196


LES LOIS

DE

L'IMITATION

ÉTUDE SOCIOLOGIQUE

PAR

G. TARDE

Membre de l'Institut

Professeur au Collège de France.

SIXIÈME ÉDITION

PARIS

FÉLIX ALCAN, ÉDITEUR

LIBRAIRIES FÉLIX ALCAN ET GUILLAUMIN RÉUNIES

108, BOULEVARD SAINT-GERMAIN, 108

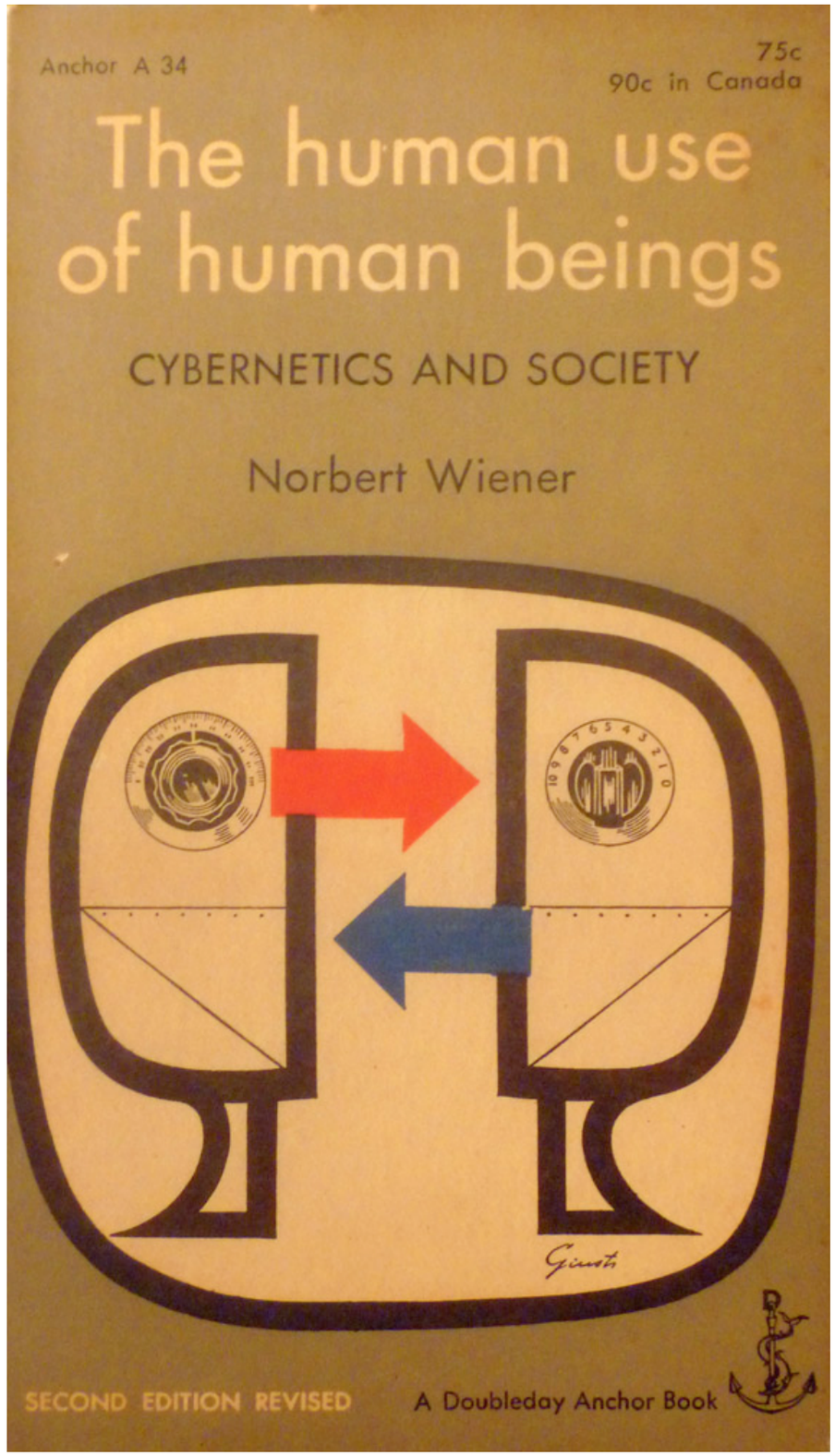
1911

Tous droits de traduction et de reproduction réservés.

Big Data in Social Sciences

“The human use of human beings”

Norbert Wiener, 1950.



Big Data in Social Sciences

“The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”

Chris Anderson, Wired, 23 juin 2008.

“The Petabyte Age is different because more is different”

“With enough data, the numbers speak for themselves”

“hypothesize, model, test is becoming obsolete”

“Petabyte allow us to say : correlation is enough”

“There is no reason to cling to our old ways. It’s time to ask : what can science learn from Google”

Big Data in Social Sciences

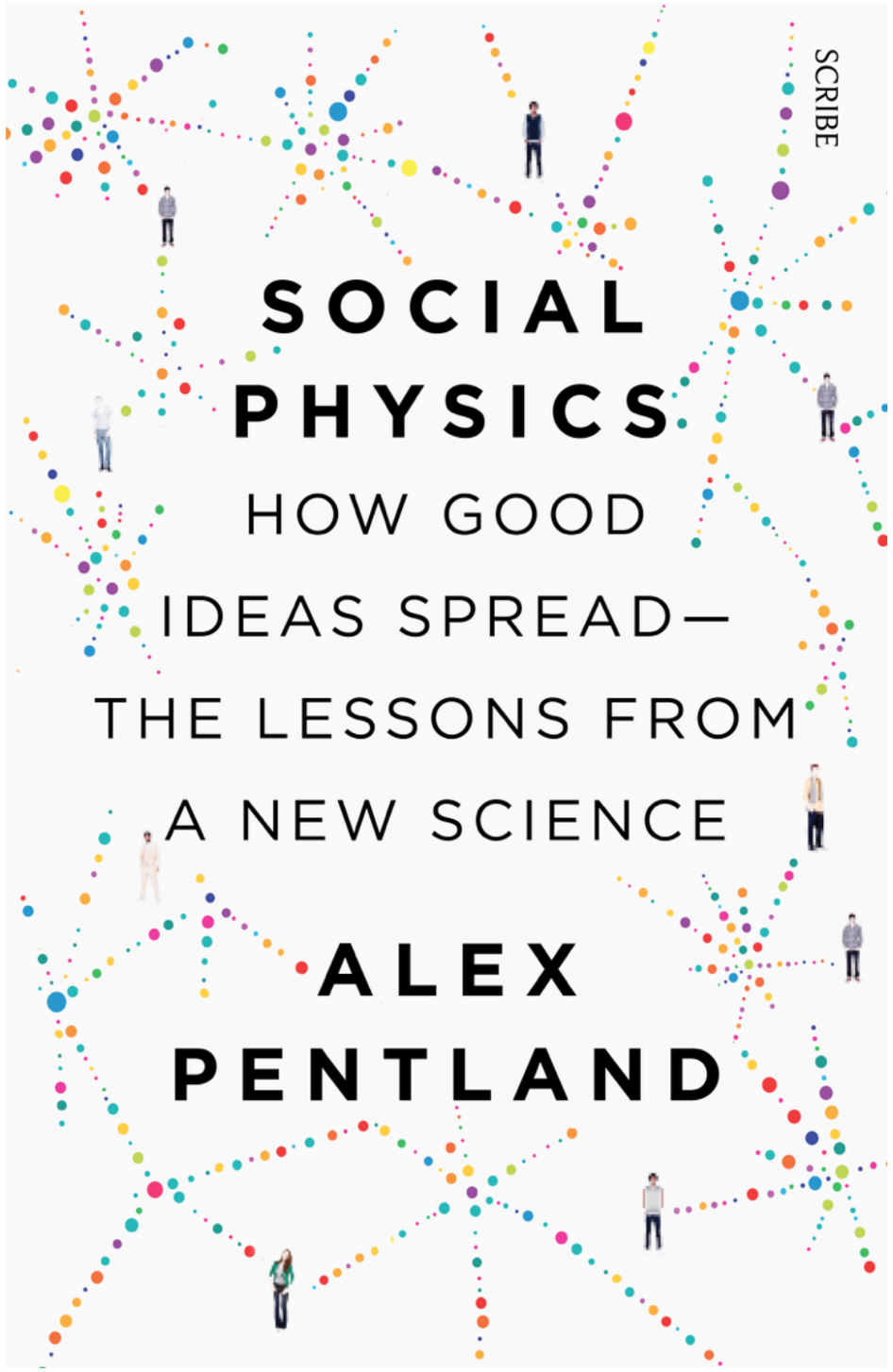
"Big Data : A revolution that will transform how we live, work, and think"

Viktor Mayer-Schönberger et Kenneth Cukier, 2013.

Big Data in Social Sciences

“Social Physics”

Alex Pentland, janvier 2014.

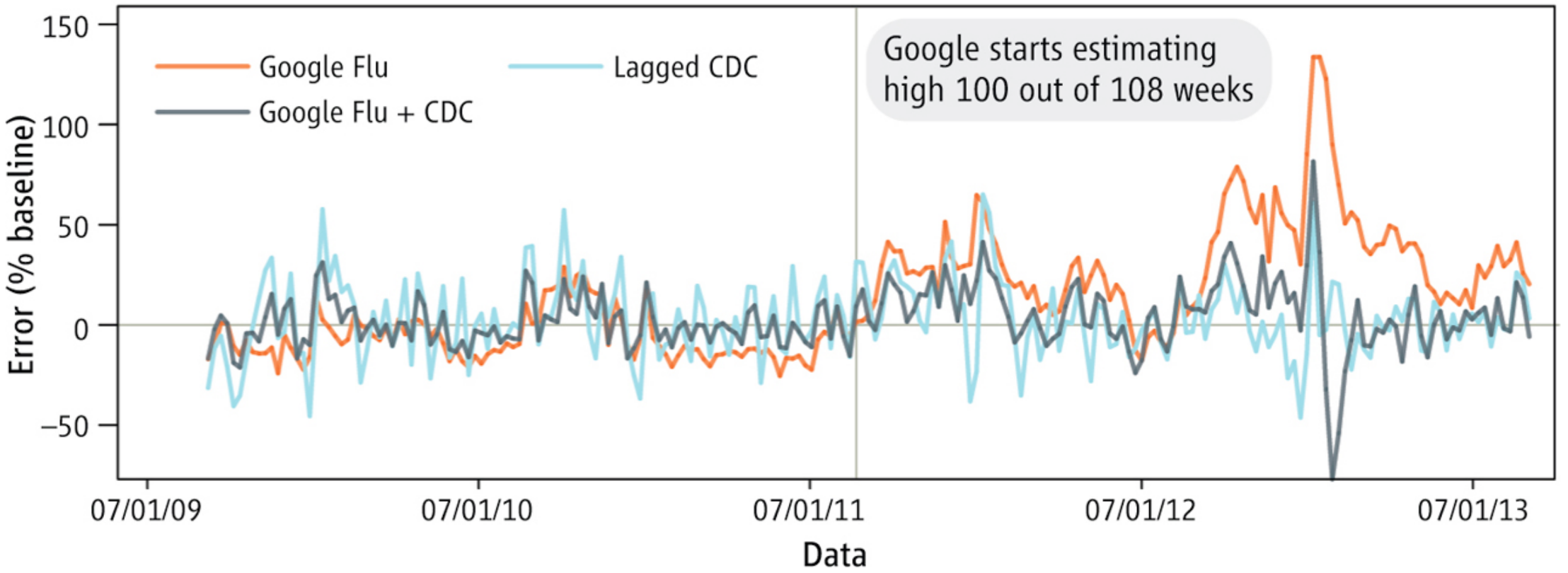


Big Data in Social Sciences

États-Unis : Données publiques sur le syndrome grippal (ILI) fournies par les Centres américains de prévention et de contrôle des maladies.

Ginsberg, Jeremy et al. 2008. “**Detecting influenza epidemics using search engine query data**.” *Nature* 457(7232): 1012–1014.

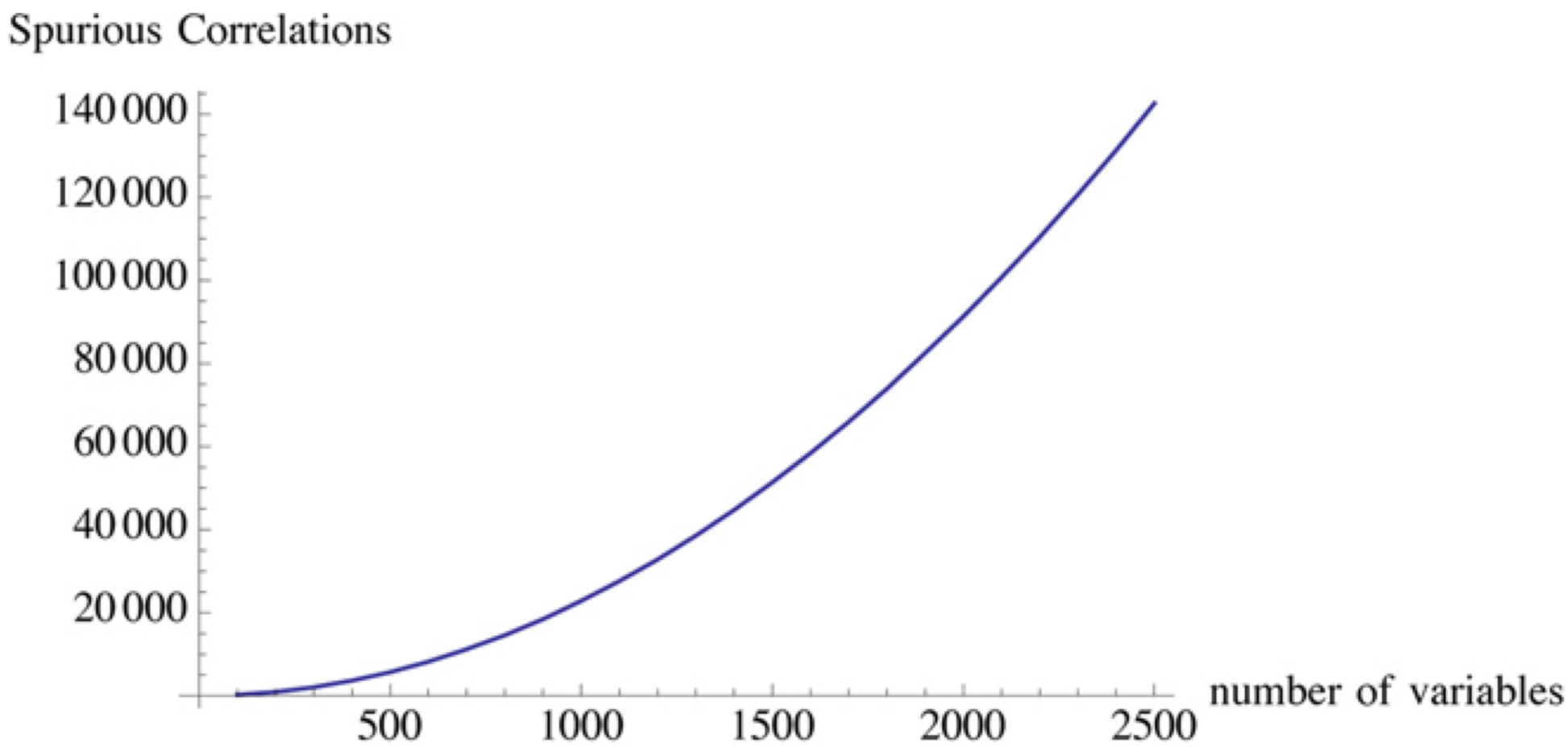
Big Data in Social Sciences



« Instead of focusing on a “big data revolution,” perhaps it is time we were focused on an “all data revolution,” where we recognize that the critical change in the world has been innovative analytics, using data from all traditional and new sources, and providing a deeper, clearer understanding of our world ».

David Lazer, et al. 2014, **The Parable of Google Flu: Traps in Big Data Analysis**, Science 343(6176): 1203-1205.

Big Data in Social Sciences

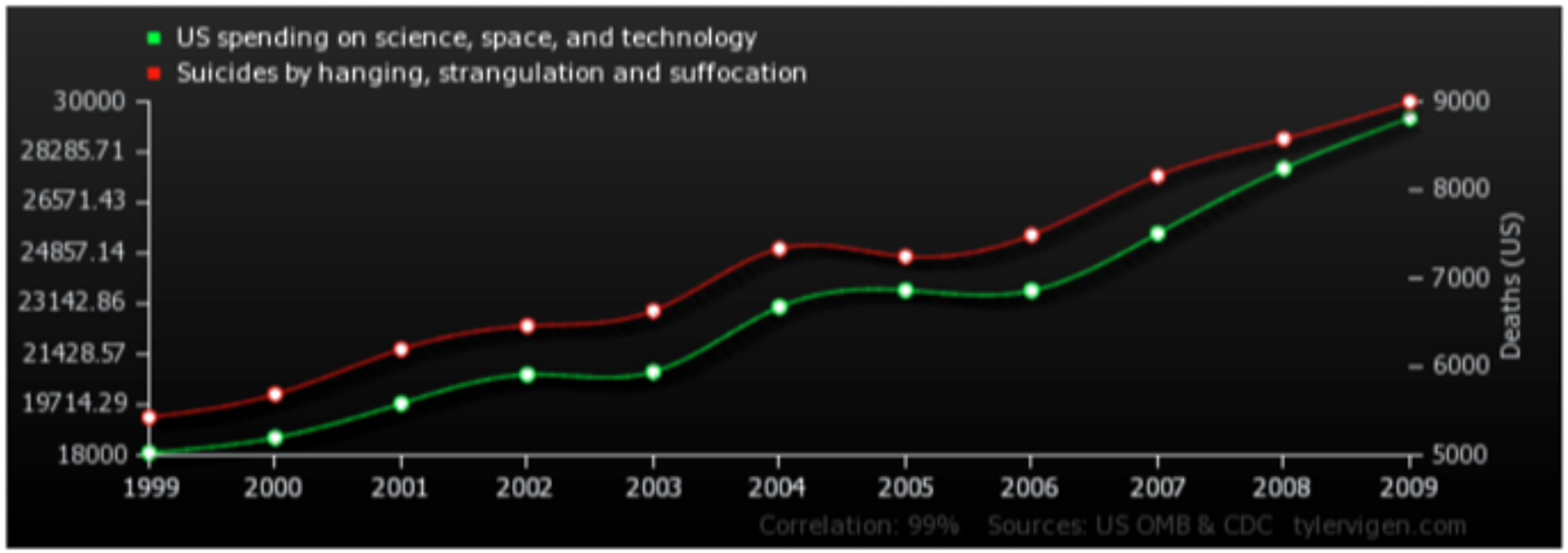


Nassim Taleb, **Antifragile**, 2014

spurious correlations

Discover a new correlation

US spending on science, space, and technology
correlates with
Suicides by hanging, strangulation and suffocation



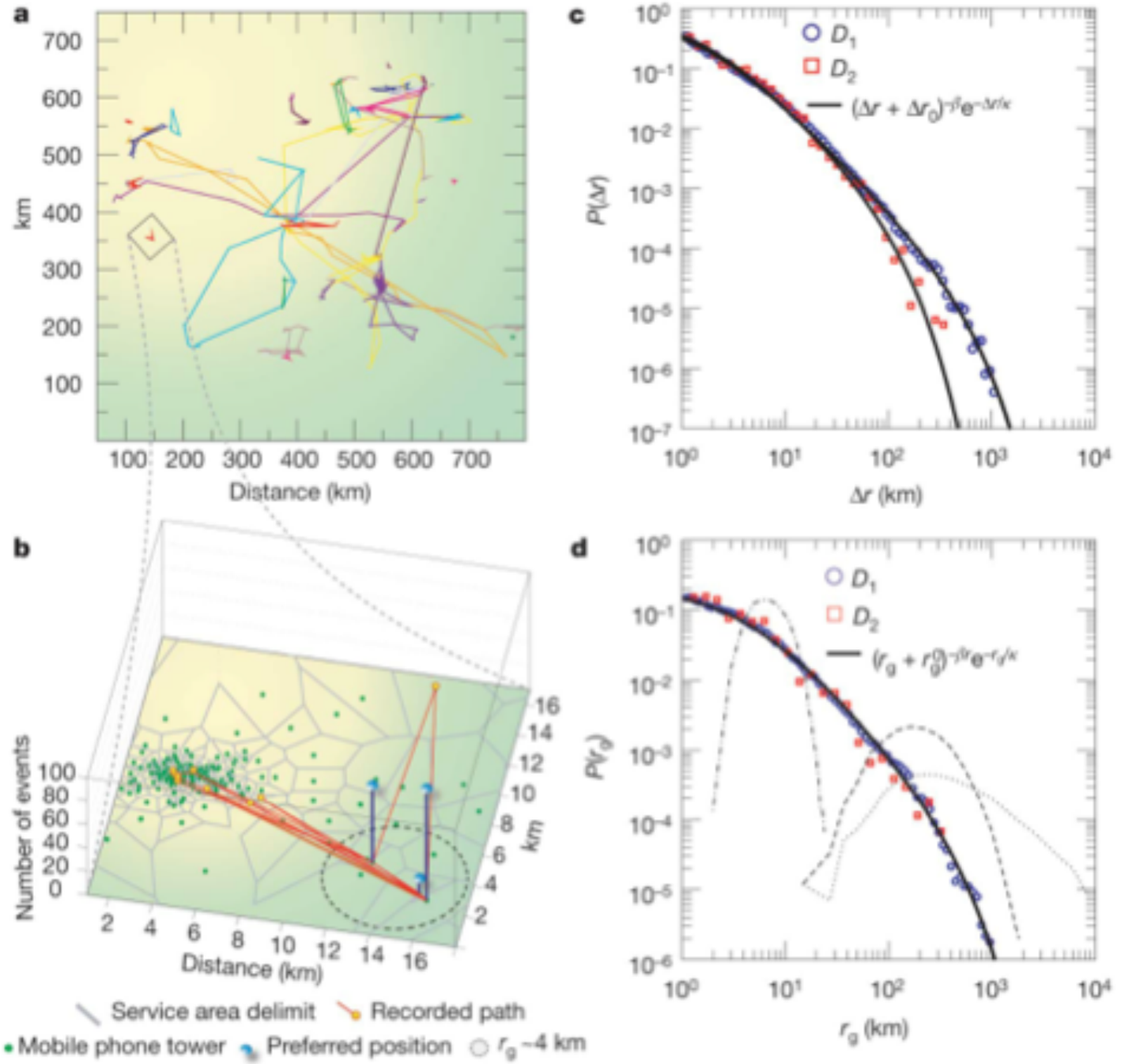
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
US spending on science, space, and technology *Millions of todays dollars (US OMB)*	18,079	18,594	19,753	20,734	20,831	23,029	23,597	23,584	25,525	27,731	29,449
Suicides by hanging, strangulation and suffocation *Deaths (US) (CDC)*	5,427	5,688	6,198	6,462	6,635	7,336	7,248	7,491	8,161	8,578	9,000

Correlation: 0.992082

Tyler Vigen, **Spurious correlations**, 2014

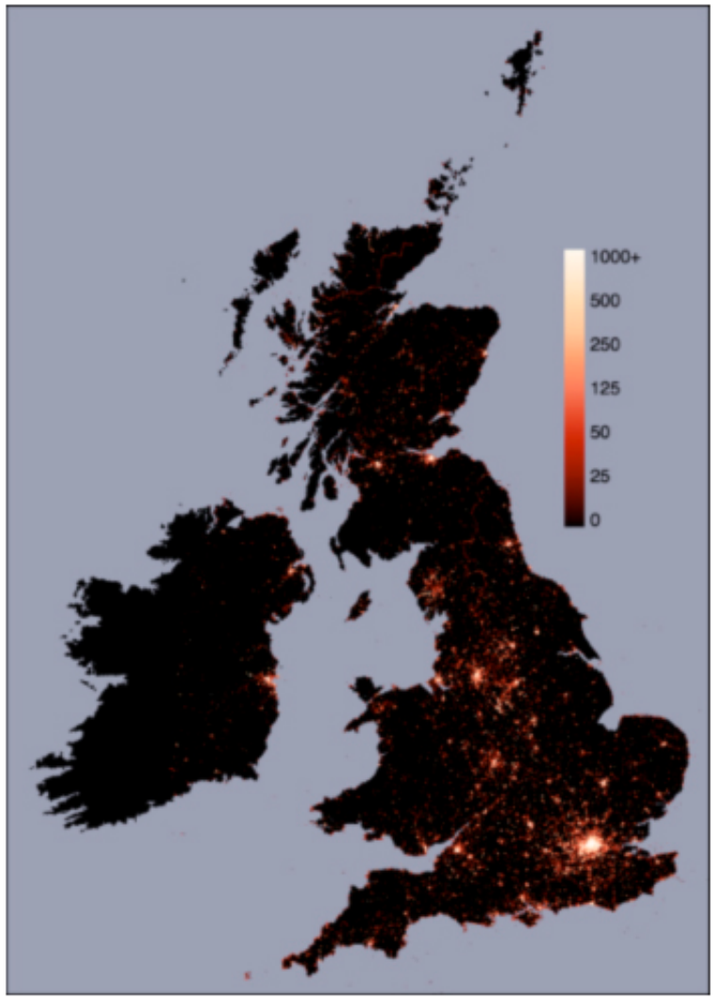
Big Data in Social Sciences

Big Data in Social Sciences

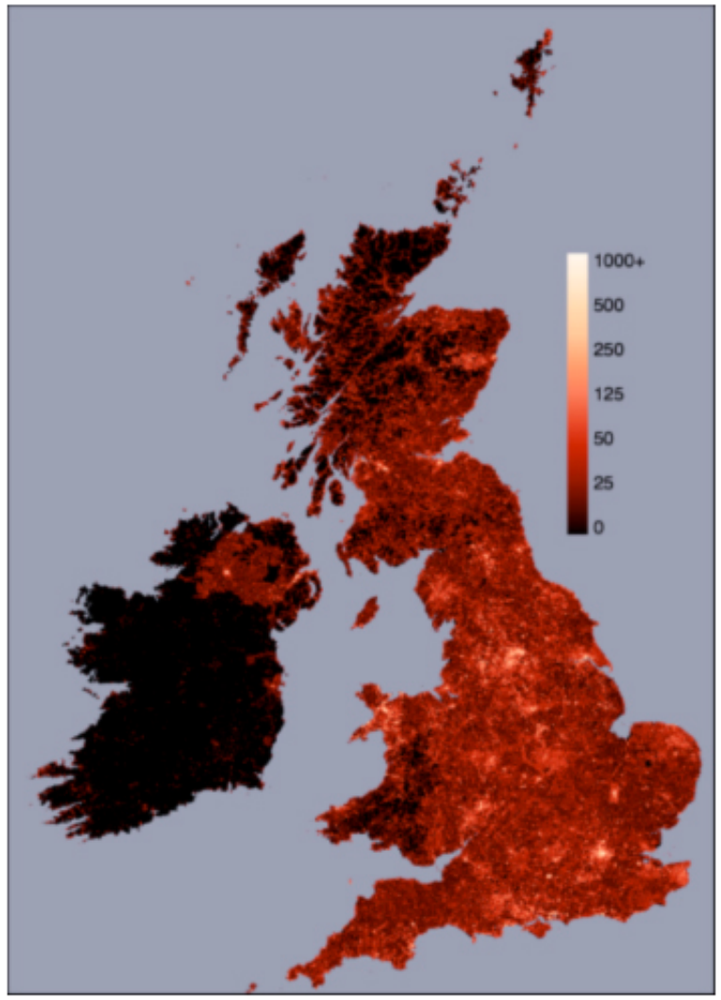


González, Marta C, César A Hidalgo, and Albert-László Barabási. 2008. "**Understanding individual human mobility patterns**." Nature 453(7196): 779–782.

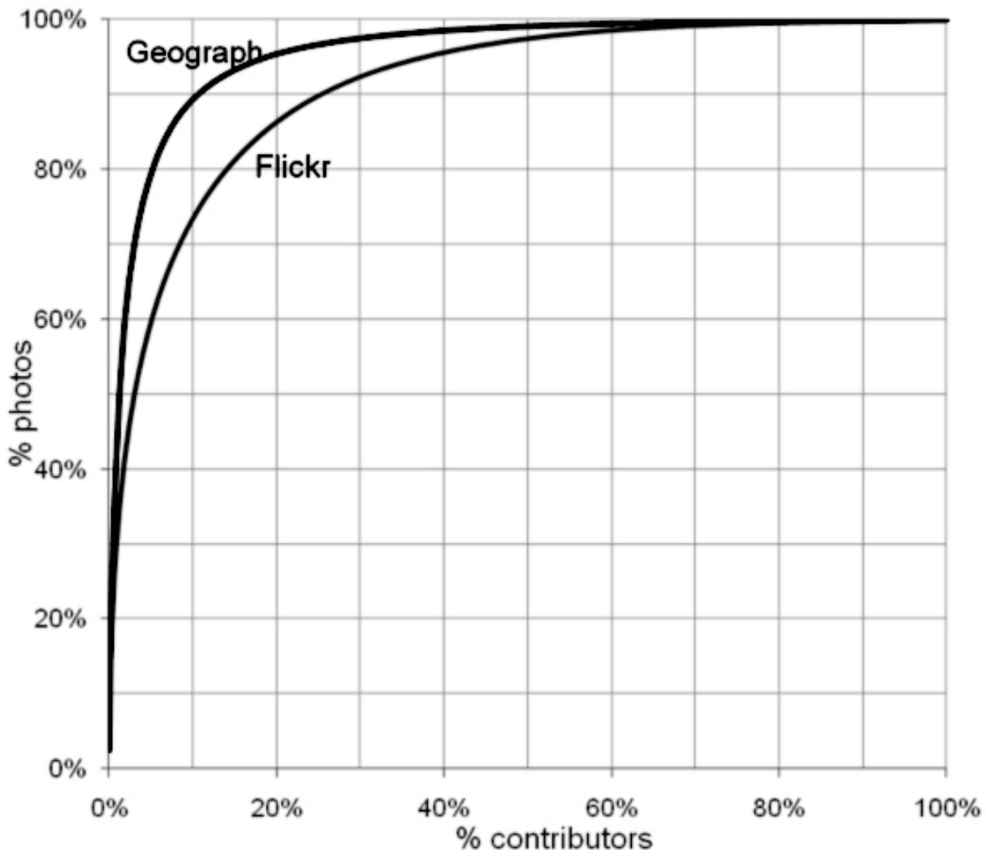
Big Data in Social Sciences



Flickr



Geograph



Purves et Edwardes, **Inégale représentativité spatiale à gauche et sociale à droite**, 2011

Big Data in Social Sciences

Open Street Map et Cloud Made, Lausanne - EPFL, 2008

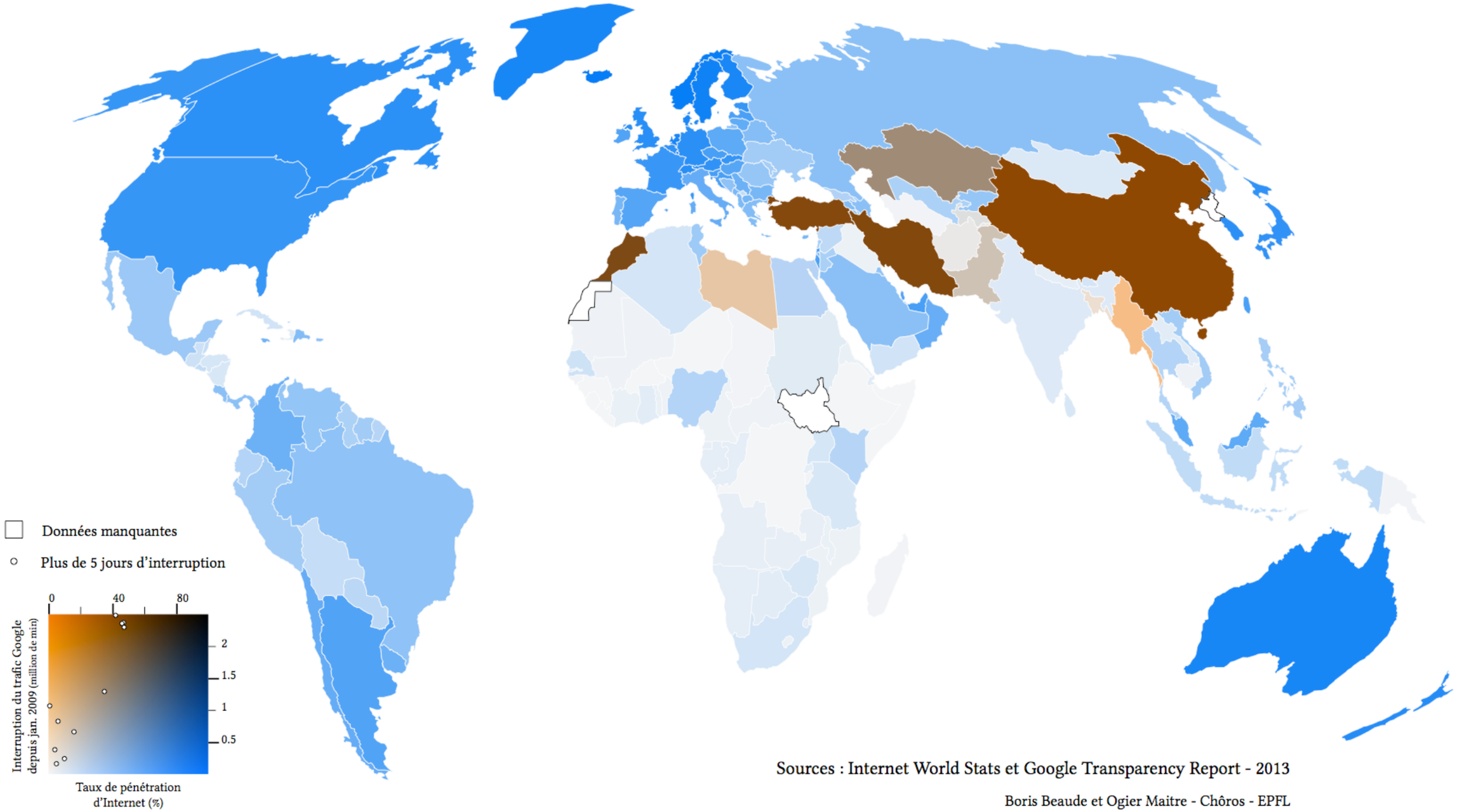
Big Data in Social Sciences

The Road to 200 Million

Facebook began as a private network for colleges and universities, but has grown into an international social networking site with almost 200 million members. Lee Byron, a member of Facebook's data team, created maps and network diagrams that show the site's expansion and use.

Lee Byron, **Facebook Growing to 200 Million**, 2008

Big Data in Social Sciences



Boris Beaude et Ogier Maitre - **L'ubiquité fracturée** - 2014

Big Data in Social Sciences

“Critical questions for Big Data”

dana boyd & Kate Crawford, Information, Communication & Society, June 2012, pp. 662–679.

« Technology is neither good nor bad; nor is it neutral »

1. Big Data changes the definition of knowledge

2. Claims to objectivity and accuracy are misleading

3. Bigger data are not always better data

4. Taken out of context, Big Data loses its meaning

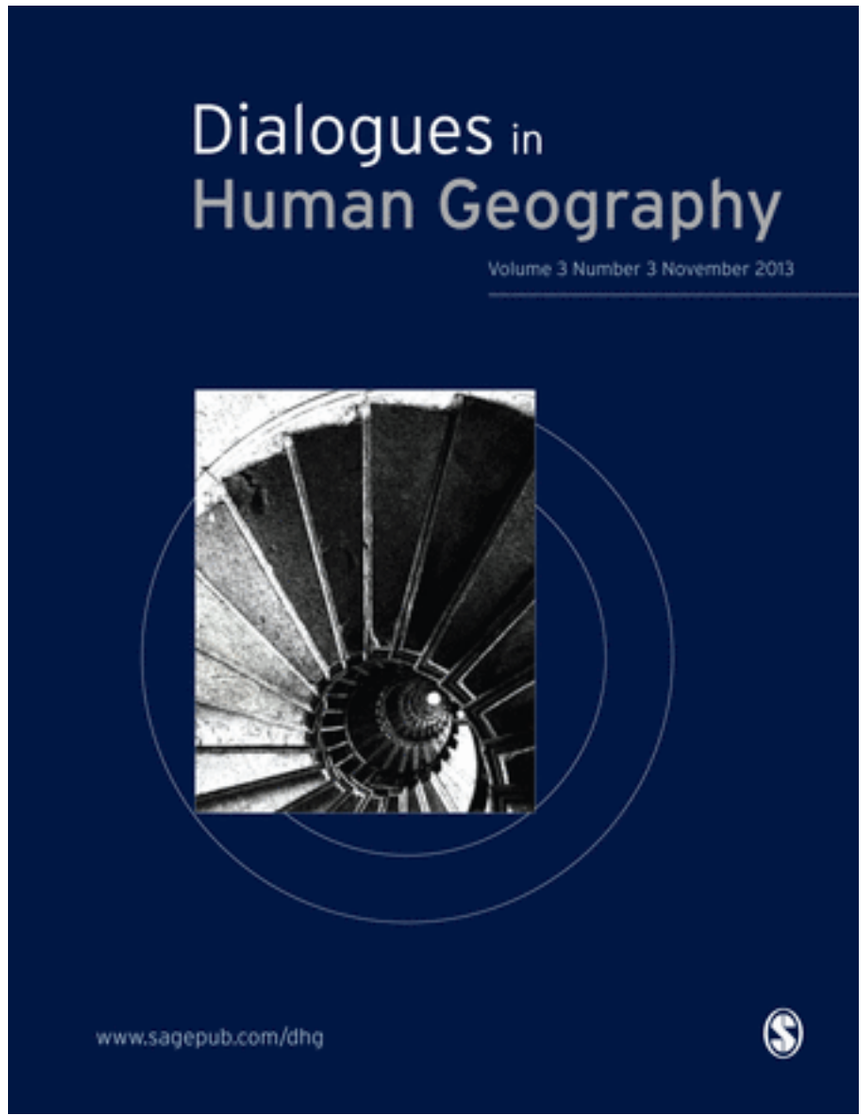
5. Just because it is accessible does not make it ethical

6. Limited access to Big Data creates new digital divides

Big Data in Social Sciences

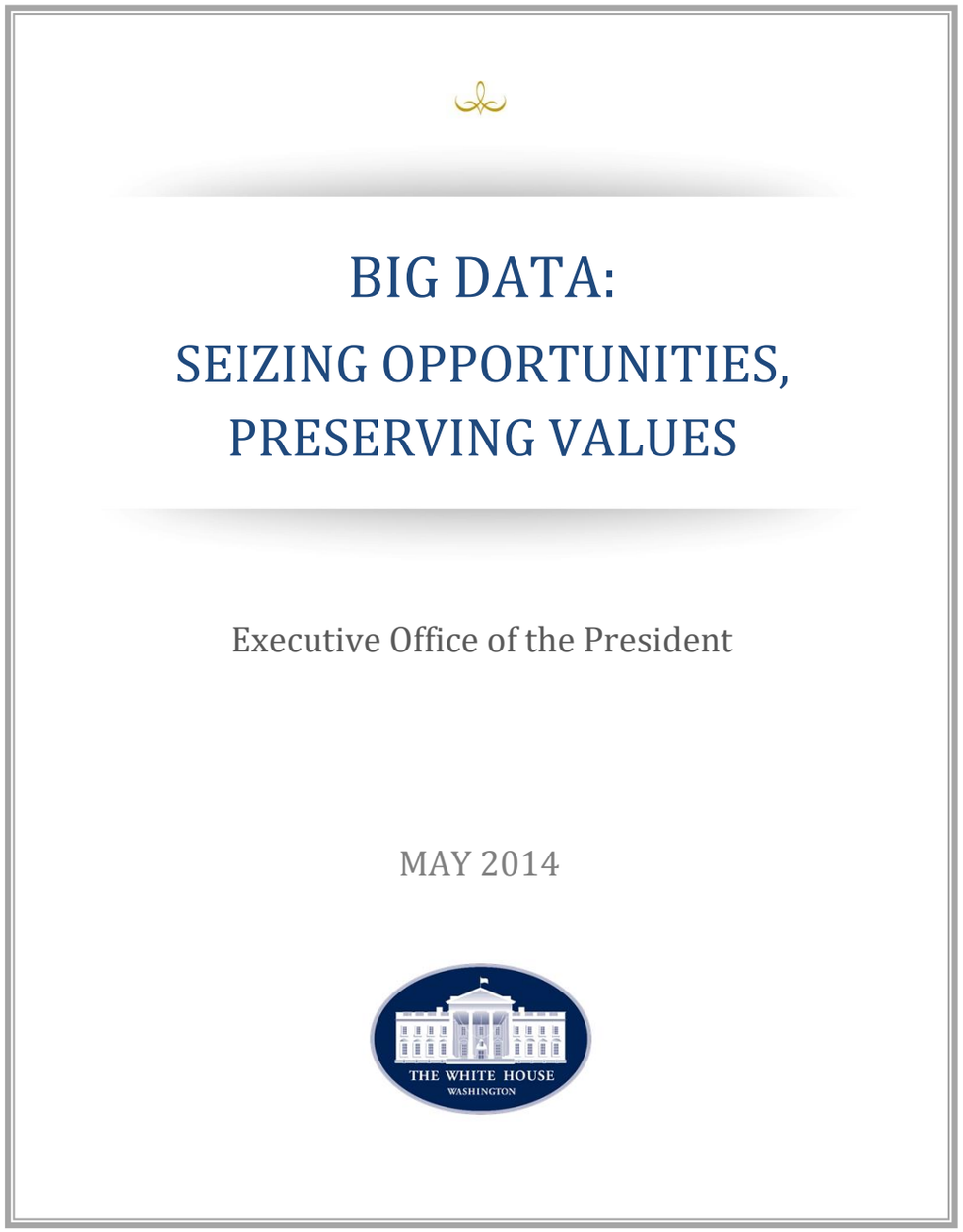
"Geography and the future of big data, big data and the future of geography"

Mark Graham and Taylor Shelton, Dialogues in Human Geography November 2013.



“Big Data : Seizing opportunities, preserving values”

Executive Office of the President, May 2014.


BIG DATA:

SEIZING OPPORTUNITIES,
PRESERVING VALUES

Executive Office of the President

MAY 2014

THE WHITE HOUSE
WASHINGTON

Big Data in Social Sciences

“Big data: are we making a big mistake?”

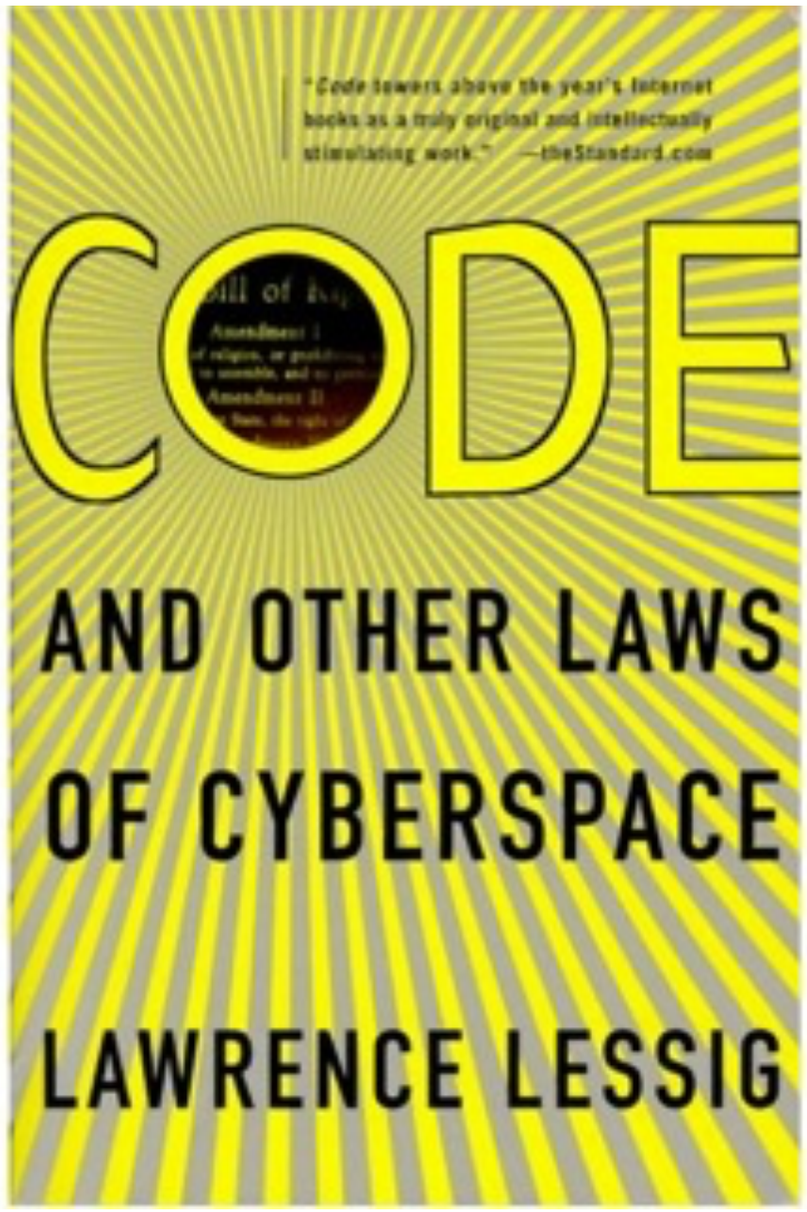
Tim Harford, Financial Times Magazine, March 28, 2014.

- numbers don’t speak for themselves
- theory free analysis is fragile
- N=all doesn’t exist
- Sampling error & Sampling biais
- Correlations are not causation
- We are not studying a stable environment

- Alfred Landon vs Franklin Roosevelt (1936) and sampling biais
- Street Bump (Boston) and sampling biais
- Target and baby clothes (2012) and false positives
- John Ioannidis, epidemiologist (2005) - Why most published research findings are false ?

“Code and other laws of cyberspace”

Lawrence Lessig, 1999.



Big Data in Social Sciences

« Les cibles de pouvoir ont glissé sur les prises de l'advenir…
C'est notre dimension de potentialité qui est visée et non notre dimension d'actualité ! »

Antoinette Rouvroy, « la politique des données personnelles », Lyon, 2013